Ho4 7-26

PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 : H04N 7/01, 7/26

A1

(11) International Publication Number: WO 97/3957

(43) International Publication Date: 23 October 1997 (23.10.9

(21) International Application Number: PCT/GB97/01034

(22) International Filing Date: 14 April 1997 (14.04.97)

(30) Priority Data:
9607645.0 12 April 1996 (12.04.96) GB

(71) Applicant *(for all designated States except US)*: SNELL & WILCOX LIMITED [GB/GB]; 6 Old Lodge Place, St. Margaret's, Twickenham, Middlesex TW1 1RQ (GB).

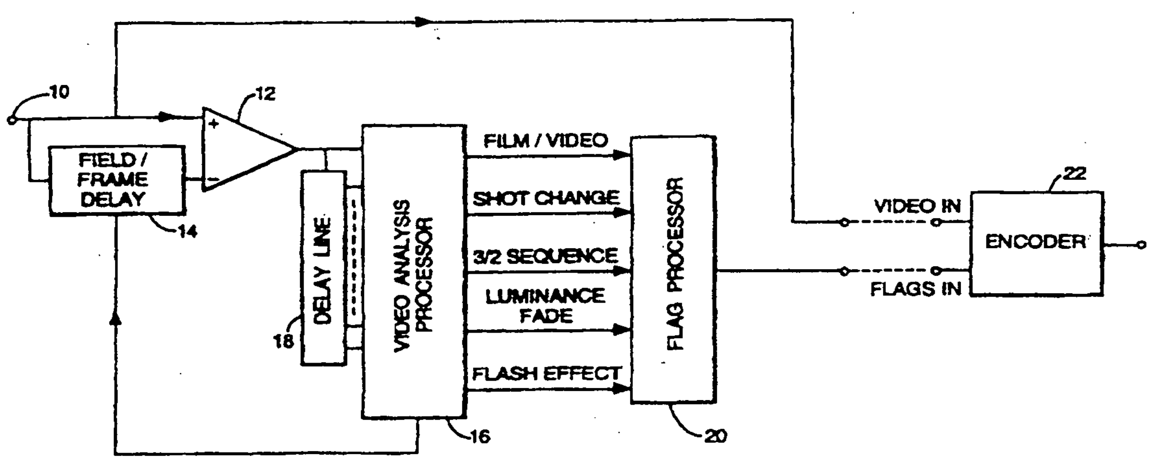
(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: KNEE, Michael, James [GB/GB]; 6 Woodbury Avenue, Petersfield, Hampshire GU32 2EE (GB). SOMERVILLE, Stuart [GB/GB]; 15 Black Cap Close, Rolands Castle, Hampshire PO9 6HD (GB).

(74) Agent: GARRATT, Peter, Douglas; Mathys & Squire, 100 Gray's Inn Road, London WC1X 8AL (GB).

(81) Designated States: AU, CA, JP, US, European patent (AT, B: CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, N! PT, SE).

Published
With international search report.

(54) Title: PROCESSING OF VIDEO SIGNALS PRIOR TO COMPRESSION



(57) Abstract

Video analysis processing upstream of a compression encoder provides for use by the encoder flags such as a shot change flag; frame/field based encoding flag; a 3:2 pull-down flag; a 25/24 telecine flag; a luminance fade flag and a flash effect flag. The encod makes use of these flags in its coding decisions to improve coding efficiency or to reduce artefacts.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PC

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece			TR	Turkey
BG	Bulgaria	HU	Hungary	ML	Mali	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MN	Mongolia	UA	Ukraine
BR	Brazil	IL	Israel	MR	Mauritania	UG	Uganda
BY	Belarus	IS	Iceland	MW	Malawi	US	United States of America
CA	Canada	IT	Italy	MX	Mexico	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NE	Niger	VN	Viet Nam
CG	Congo	KE	Kenya	NL	Netherlands	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NO	Norway	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	NZ	New Zealand		
CM	Cameroon			PL	Poland		
CN	China	KR	Republic of Korea	PT	Portugal		
CU	Cuba	KZ	Kazakstan	RO	Romania		
CZ	Czech Republic	LC	Saint Lucia	RU	Russian Federation		
DE	Germany	LI	Liechtenstein	SD	Sudan		
DK	Denmark	LK	Sri Lanka	SE	Sweden		
EE	Estonia	LR	Liberia	SG	Singapore		

09/806091
JC10 Rec'd PCT/PTO 26 MAR 2001

PROCESSING OF VIDEO SIGNALS PRIOR TO COMPRESSION

This invention relates to the processing of video signals prior to MPEG encoding or other compression processes.

Compression schemes such as MPEG2 work optimally with ordered sequences of input pictures having the relatively high level of correlation – one to the next – that is characteristic of "true" unedited video. Shot changes or the results of telecine operation and edits between true and telecined video, disturb these ordered sequences and can degrade the performance of an encoder.

It is an object of the present invention to provide for processing of video signals prior to encoding, so as to provide information capable of enhancing the compression or at least minimising degradation.

Accordingly, the present invention consists in one aspect in a video signal processing for use upstream of a compression encoder and adapted to provide for use by the encoder one or more flags from the group consisting of a shot change flag; a frame/field based encoding flag; a 3:2 pull–down flag; a 25/24 telecine flag; a luminance fade flag and a flash effect flag.

In one form of the invention, a shot change flag is provided to the encoder for the insertion, if buffer occupancy permits, of an I picture. If the buffer occupancy does not permit, the shot change flag may trigger pre–filtering of the I picture so as to reduce the encoder buffer requirement or encoding as a P picture.

Advantageously, the shot change flag is provided in advance so as to enable the encoder to encode pictures preceding the shot change in a manner so as to provide sufficient buffer occupancy for an I picture to be inserted at the shot change.

In another form of the invention a frame–field based encoding flag is provided to enable the encoder to employ a frame based encoding wherever possible.

In another form of the invention, a 3:2 pull–down flag is capable of

indicating the 3:2 sequence whether or not interrupted by edit.

In yet a further form a 25/24 telecine flag will identify locations where fields making up a film frame are straddled over a video frame.

In still a further form of the invention, a luminance fade flag will identify for the encoder a luminance fade in the encoder to make use of correlation between pictures not withstanding progressive changes in luminance.

In still a further form of the invention, a flash effect flag serves to look at histogrammed luminance intensities across a number of pictures to identify sudden luminance changes and to enable the encoder to avoid degradation of the encoded sequence.

The invention will now be described with reference to the accompanying drawings in which:-

Figure 1 is a block diagram showing apparatus according to one embodiment of the present invention;

Figure 2 is a diagram illustrating the derivation of a shot change flag in film material;

Figure 3 is a diagram illustrating the derivation of a shot change flag in video material;

Figure 4 is a diagram illustrating the derivation of a frame/field flag;

Figure 5 is a diagram illustrating the derivation of a 3:2 pull-down flag;

Figure 6 is a diagram similar to Figure 5 but includes a sequence discontinuity;

Figure 7 is a diagram illustrating the derivation of a 625 frame pairing

flag; and

Figure 8 is a diagram illustrating the derivation of a luminance fade flag.

Referring initially to Figure 1, there is shown a comparator 12 receiving the input video signal from input terminal 10 and a field or frame delayed signal via field/frame delay 14. A video analysis processor 16 receives the current field/frame difference and – via delay line 18 – an appropriate number of past differences and serves to generate pre–processing flags in a manner to be described. The processor 16 provides a control input to the field/frame delay 14. The flags are (if necessary) converted into the appropriate form for transmission in flag processor 20 and are made available to a downstream compression encoder 22, alongside the input video signal.

There are a number of different input formats which are identified by the video analysis processor.

- Film originated 525 (using 3:2 pull–down technique)
- Film originated 625 (telecine operates at 25/24 rate)
- Video originated 525\625 (2:1 interlaced)
- Any electronically post produced combination of the above

The presence of a picture or shot change is detected by the analysis of picture difference between frames or fields. Several correlation techniques are available such as integrated low pass filtered luminance and chrominance differences or correlation of histogrammed luminance intensities.

Film material which is frame based will generate a strong intra–frame

correlation i.e. there will be no motion differences between fields which originate from the same film frame.

On the other hand, video originated material which has individual fields representing different points in time has no frame correlation. Nevertheless, at the presence of a shot change the integrated field and frame difference outputs will show a distinctive pattern which can be correlated to identify the exact field where the shot change occurs. This is illustrated for the case of film material in Figure 2 and for the case of video in Figure 3.

The processor 16, which can take the form of a dedicated microprocessor, is used to monitor a number of previous field and frame differences as well as other parameters which may also be taken into account, such as noise floor level.

The shot change flag is a single line active high immediately prior to the shot change and is cleared at the end of the following field. Therefore the line is active for a period of one field and can be used to influence a downstream encoder in such a manner that it can reduce the visibility of the temporal discontinuity caused by a scene change.

The flag processor can also be programmed to supply the flag in advance of the shot change by a number of fields, this parameter being user programmable in field steps, typically from one to six fields.

This flag is provided to overcome an MPEG encoder trying to predict across a cut boundary. The intention is to modify the Group of Pictures (GOP) structure to produce the best possible result. The constraints on this modification are the buffer occupancy at the point of the edit. If the buffer will allow an I picture to be inserted at the beginning of the sequence after the edit point this is the ideal situation. However, this might not be possible so the following options could be considered:-

i) encoding as a P picture so that the prediction of the pictures leading up to the edit point are only based on the preceding I picture, or

ii) pre-filtering the I picture so that the buffer requirement can be reduced when the picture is subsequently encoded.

The idea of providing shot change information can be further refined by using advanced shot change information. This would allow pictures leading up to the shot change to be encoded so that the necessary buffer space can be obtained to accommodate the shot change by inserting an I picture. For example, an I picture which would otherwise have been encoded in the period now known to precede a shot change, can be deferred until the location of the shot change.

Film originated material has a high level of picture correlation on fields which are extracted from the same film frame. A film originated signal will therefore produce a frame rate output signal which can be identified and used to identify the source as film originated. A video originated source however, has no such frame correlation.

A single bit is used to indicate film/video origination and is provided at the start of each field. If the material contains mixed film and video originated material then the flag changes state immediately prior to the first field of the film/video edit. This is illustrated in schematically in Figure 4, which shows the frame/filed flag remaining high only in the presence of the characteristic film signature.

Ideally an encoder wishes to encode using the most efficient coding modes. Therefore the use of frame based encoding as opposed to field based encoding is preferable. So a flag to indicate the nature of the input video is highly desirable to ensure that the quality of the video is preserved whilst coding efficiency is maintained.

Film material which is scanned by telecine to produce a 525 line output uses the well known 3:2 pull-down technique to insert an additional field for each pair of film frames. This has the effect of increasing the output field rate to 60Hz required for 525 distribution.

The output video signal has a distinctive pattern of 2 fields followed by 3 fields followed by 2 fields etc.. This pattern can be reliably detected using techniques described earlier such as integrated low pass filtered luminance between frames. The presence of the repeat field can be used to identify the noise floor level since the frame difference output will produce a very low

level for one field in every five. This occurs whenever a film frame has been scanned to produce three output fields of which two are identical and separated by one field. In this case the residual noise value is comprised of film grain noise, quantisation noise etc..

The five field sequence can be identified by a specific code for each field in the sequence.

This is achieved using a 3 bit code running from 0 to 4 as shown in Figure 5.

The removal of the repeat field of the 3/2 sequence greatly increases coding efficiency. To do this well, temporal analysis of the material is required and for a description of an appropriate technique, reference is directed to US-A- 5,255,091. What should be stressed is important is not only the detection of the 3/2 sequence, but also the ability to detect an interrupted 3/2 sequence as is experienced in material that has been edited as film and then transcribed to video. This changing of the 3/2 sequence and being able to not only detect this but also convey it to the compression encoder makes this approach very powerful.

Although the process of detecting the 3:2 sequence is relatively straight forward, complications arise whenever the 525 material is electronically edited. This causes temporary discontinuities in the 3:2 sequence at the edit point. Downstream equipment which relies on the accuracy of this flag could be affected if the sequence counter does not correspond exactly to the video at the edit point. This embodiment uses a correlation algorithm to:-

- detect the edit
- analyse the 3:2 sequence following the edit
- provide accurate co-timed sequence count at the output

An example of an electronic edit to a film originated master is shown

in Figure 6.

In this case, the edit has caused a discontinuity in the sequence count resulting in the counter value changing from a 2 back to a 1 at the edit point. This embodiment will accurately detect the edit and adjust the sequence count value until it is co–timed with the video output.

Film source material is scanned directly to produce a 625 line 50Hz output by running the telecine at 25/24 normal speed. The resulting output has pairs of fields corresponding to each original film frame. No additional fields are required as the resulting frame rate of 25Hz is exactly correct for the 625 line standard. The frame pairs are identified by analysis of field differences. A characteristic frame rate signal is produced by the frame pairing which can be extracted and provided to downstream equipment as shown in Figure 7.

Normally, the two fields making up a film frame will be conveyed as the two fields of a video frame. However, there are situations when this relationship does not hold and the two fields making up a film frame are straddled over a video frame. The problem with this splitting up of the "film field pairing" is that if an encoder tries to use the most efficient frame coding modes the field pairing is incorrect and consequently the quality of the encoded pictures is poor and/or the encoding efficiency is less than optimal. Flagging this situation allows corrective action to be taken so the correct relationship between the field and frame pairing can be maintained. This corrective action might take the form of the introduction of a field delay. Therefore the more coding efficient frame based encoding can be used. If it is not possible to take this "corrective action" the flagging of this situation means that the encoder will not try and use frame based techniques if it is not correct to do so.

The mixer is a standard electronic edit tool which is widely used to control switching between two sources. Typically described as a dissolve or fade, one input is reduced in amplitude while a second input is increased by a corresponding amount. The gradient signal is usually linear on each signal. The linear gradient is usually applied to each field and may affect

many fields. On film originated material the mix can mask the normal frame correlation of field pairs since mix can cause field-to-field differences. A picture correlation will have a minimum value corresponding to the mid point of the mixing function. The microprocessor can use this information along with the field and frame difference outputs to provide a co-timed flag indicating the duration of the video mix as shown in Figure 8.

In the presence of a fade to or from black or a cross-fade, many motion estimators within the compression encoder have difficulty in producing the correct motion vectors. By providing a flag to indicate that these fades are occurring, the motion estimator can then take into account the luminance changes experienced during a fade and so improve the quality of the motion vectors produced. The type of action that would be taken during a fade would be to adjust the thresholds used between the search window and the trial block so that the luminance change can also be included in this comparison. Luminance fades and cross-fades can be thought of as linear or non-linear keying. The luminance fade/cross-fade information can be presented as a linear key but provided the encoder had been provided with the transfer function of the key involved, a non-linear key could also be used.

In compression encoders having motion estimators which remain effective in the presence of luminance fades, such as those utilising phase correlation, the luminance fade/cross -fade information may nonetheless remain of value. In the case of a linear fade, for example, it may prove beneficial to force the GOP structure IBPBPBI, that is to say to employ single B pictures between reference pictures.

The video analysis processor can generate a flash effect flag by looking at histogrammed luminance intensities across a number of pictures to identify sudden luminance changes. The encoder may make use of this flag to ensure that a picture which suffers from a flash effect is not used as a reference picture. In other words, the encoder may react to a flash effect flag by forcing the coding of a B picture. In this way, the encoder can avoid or reduce the degradation of the encoded sequence that would normally accompany a photographic lighting or other flash effect.

CLAIMS

1. A method of video signal processing for use upstream of a compression encoder and adapted to provide for use by the encoder one or more flags from the group consisting of a shot change flag; a frame/field based encoding flag; a 3:2 pull-down flag; a 25/24 telecine flag; a luminance fade flag and a flash effect flag.

2. A method according to Claim 1, wherein the 3:2 pull-down flag is capable of indicating the 3:2 sequence whether or not interrupted by edit.

3. A method according to Claim 1, wherein the 25/24 telecine flag is capable of indicating the locations where fields making up a film frame are straddled over a video frame.

4. A method of video signal processing utilising a video analysis processor upstream of a compression encoder, wherein the video analysis processor provides a shot change flag on detection of a shot change in the video signal and wherein the encoder, if buffer occupancy permits, inserts an intra picture coded picture (I picture) on receipt of a shot change flag.

5. A method according to Claim 4, wherein, if the buffer occupancy does not permit the insertion of an I picture, the shot change flag triggers pre-filtering of the I picture so as to reduce the encoder buffer requirement for encoding as a P picture.

6. A method according to Claim 4, wherein, the shot change flag is provided in advance so as to enable the encoder to encode pictures preceding the shot change in a manner so as to provide sufficient buffer occupancy for an I picture to be inserted at the shot change.

7. A method according to Claim 6, wherein pictures preceding the shot

change are encoded without use of I pictures so as to enable an I picture to be inserted at the shot change

8. A method of video signal processing utilising a video analysis processor upstream of a compression encoder, wherein the video analysis processor provides a frame–field based encoding flag and wherein the encoder receives said flag and employs a frame based encoding wherever possible.

9. A method of video signal processing utilising a video analysis processor upstream of a compression encoder, wherein the video analysis processor provides a luminance fade flag and wherein the encoder serves to make use of correlation between pictures not withstanding progressive changes in luminance.

10. A method of video signal processing utilising a video analysis processor upstream of a compression encoder, wherein the video analysis processor serves to look at histogrammed luminance intensities across a number of pictures to identify sudden luminance changes and thereby generate a flash effect flag and wherein the encoder serves to make use of the flash effect flag to avoid degradation of the encoded sequence.

11. A method according to Claim 10, wherein the compression encoder employs the flash effect flag to ensure that pictures suffering from flash effect are not coded as reference pictures.

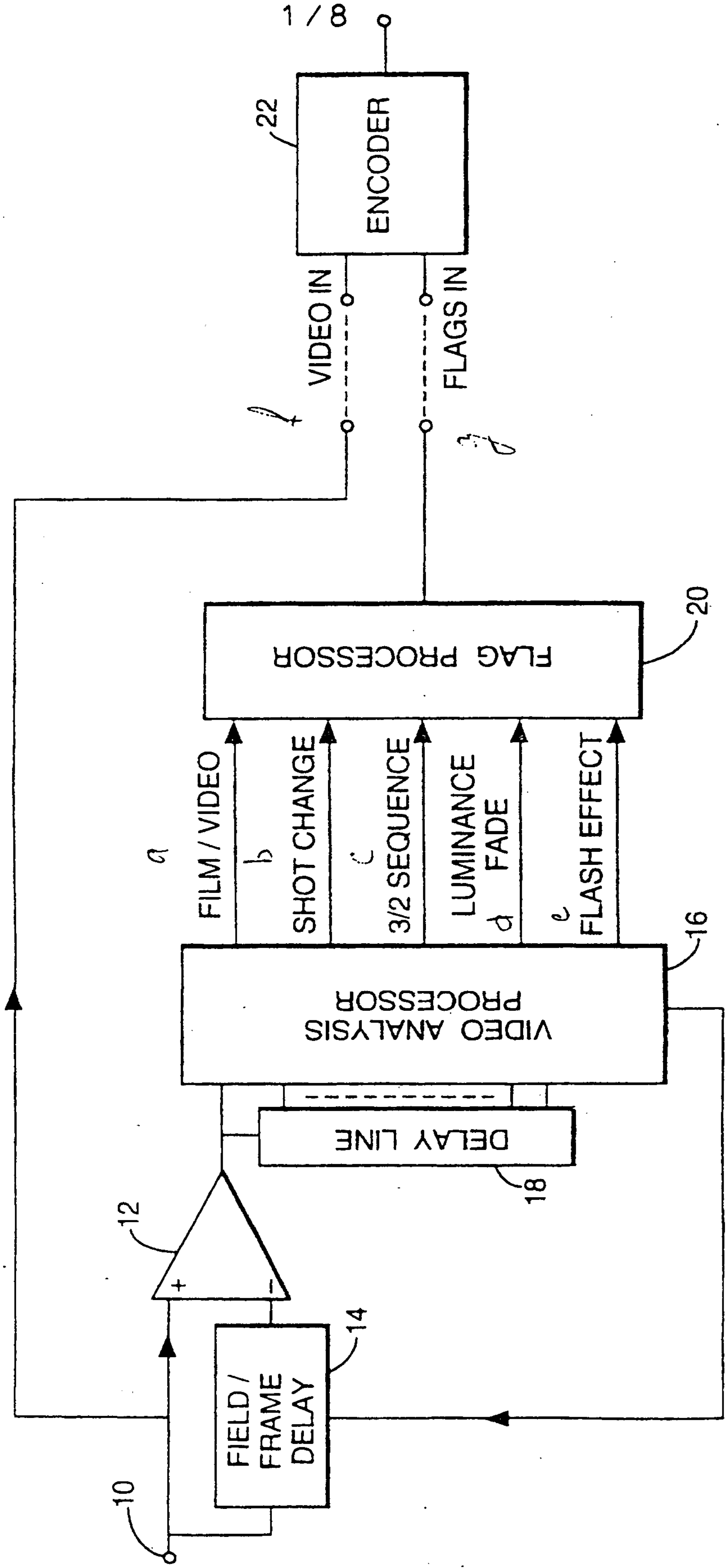
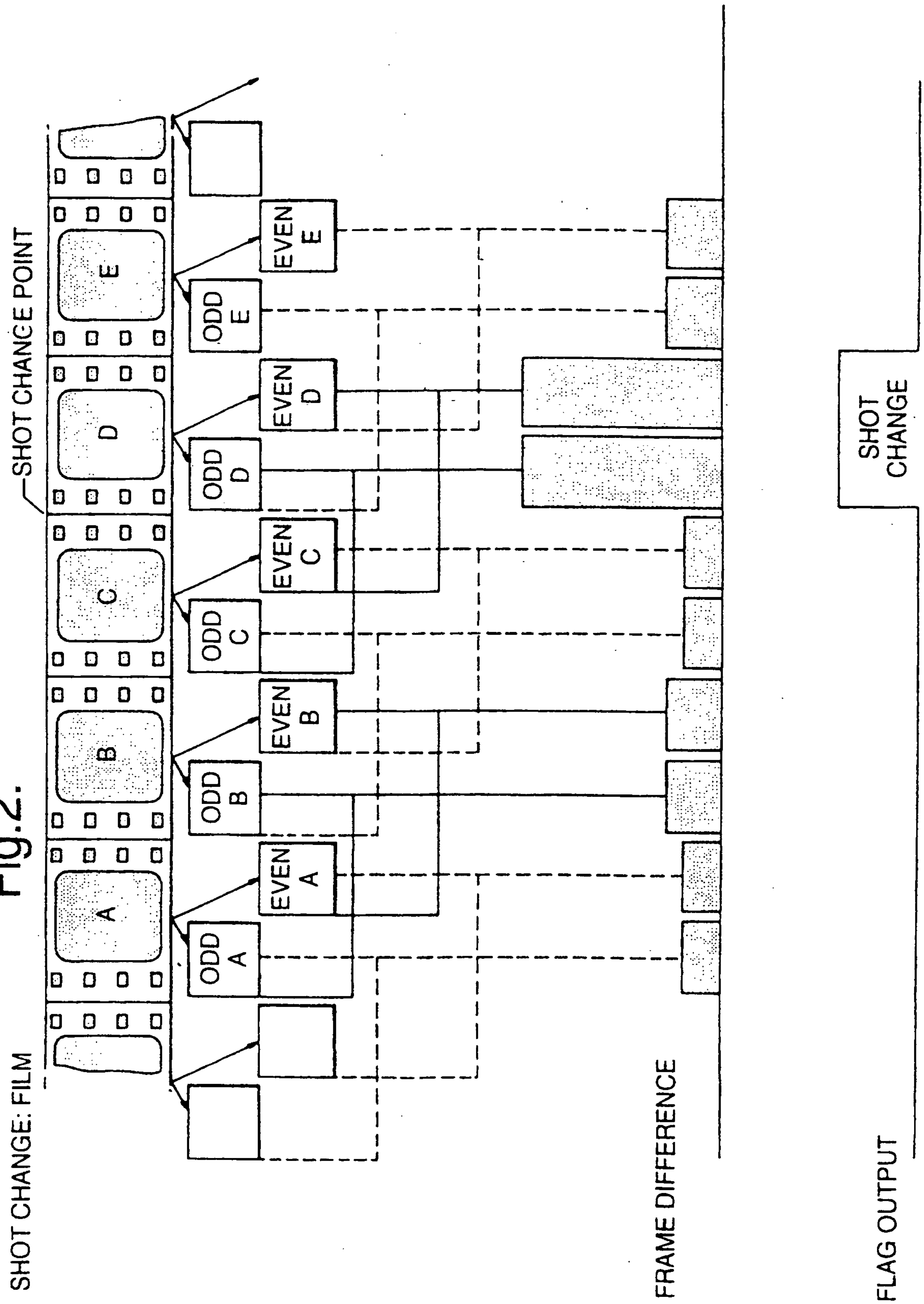

ENCODER
22
VIDEO IN
FLAGS IN
f
g
FLAG PROCESSOR
20
FILM / VIDEO
SHOT CHANGE
3/2 SEQUENCE
LUMINANCE FADE
FLASH EFFECT
a
b
c
d
e
VIDEO ANALYSIS PROCESSOR
16
DELAY LINE
18
12
+
−
FIELD / FRAME DELAY
14
10

Fig.1.

Fig.2.


SHOT CHANGE: FILM
SHOT CHANGE POINT
A
B
C
D
E
ODD A
EVEN A
ODD B
EVEN B
ODD C
EVEN C
ODD D
EVEN D
ODD E
EVEN E
FRAME DIFFERENCE
FLAG OUTPUT
SHOT CHANGE

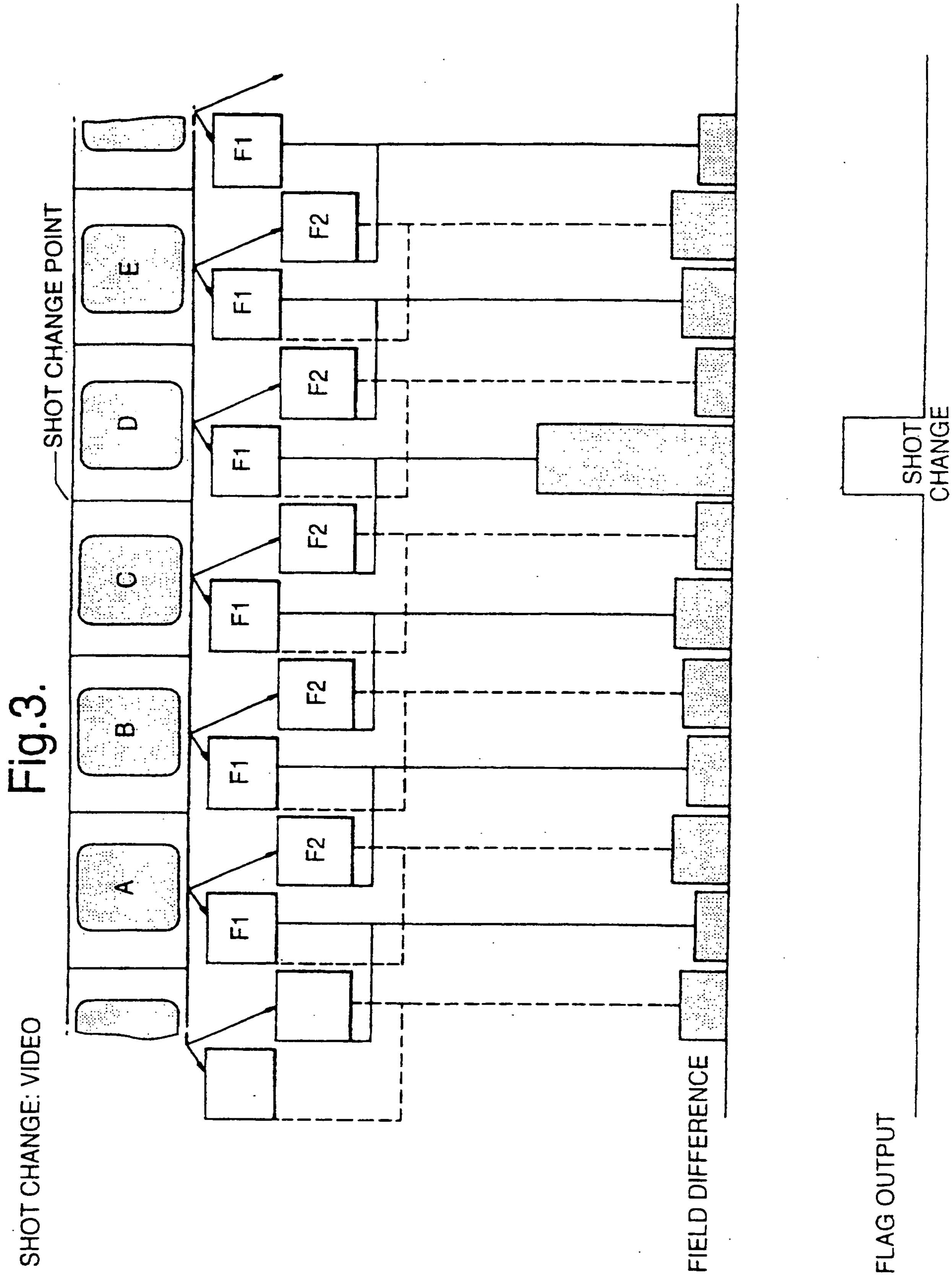


Fig.3.
SHOT CHANGE: VIDEO
SHOT CHANGE POINT
A
B
C
D
E
F1
F2
FIELD DIFFERENCE
FLAG OUTPUT
SHOT CHANGE

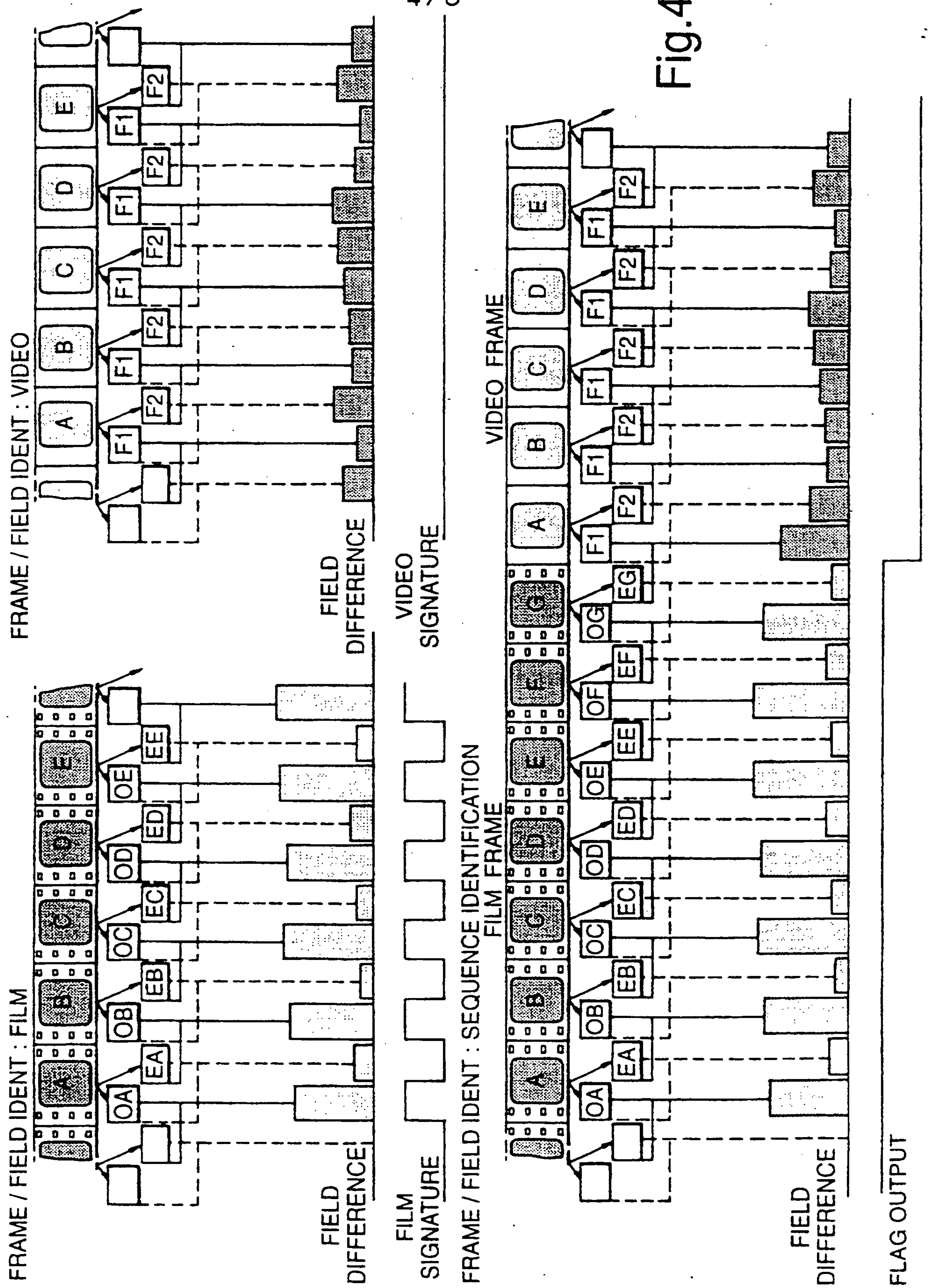

FRAME / FIELD IDENT : FILM
A B C D E
OA EA OB EB OC EC OD ED OE EE
FIELD DIFFERENCE
FILM SIGNATURE
FRAME / FIELD IDENT : VIDEO
A B C D E
F1 F2
FIELD DIFFERENCE
VIDEO SIGNATURE
FRAME / FIELD IDENT : SEQUENCE IDENTIFICATION
FILM FRAME
VIDEO FRAME
A B C D E G
OA EA OB EB OC EC OD ED OE EE OF EF OG EG
F1 F2
FIELD DIFFERENCE
FLAG OUTPUT

Fig.4.

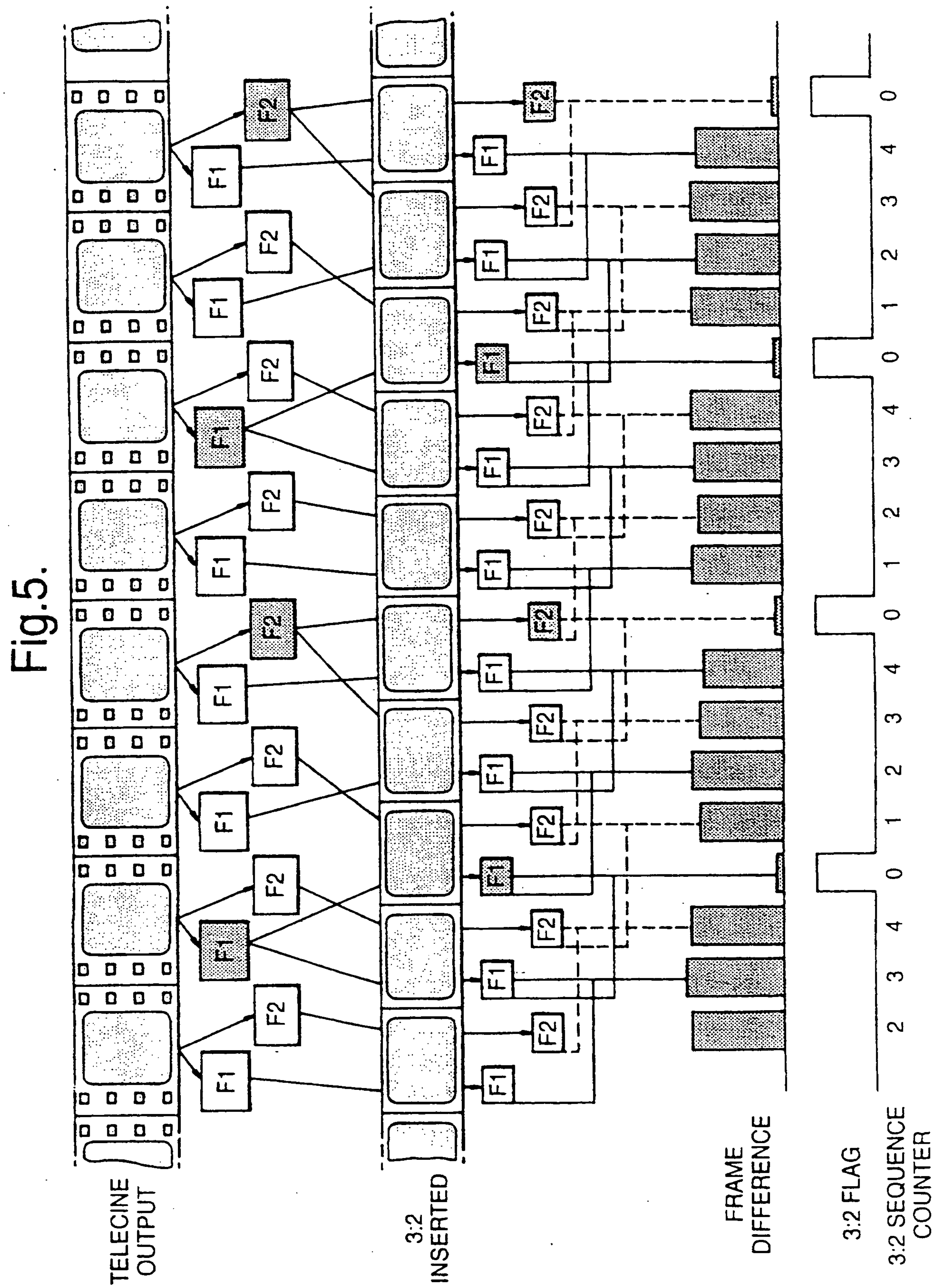
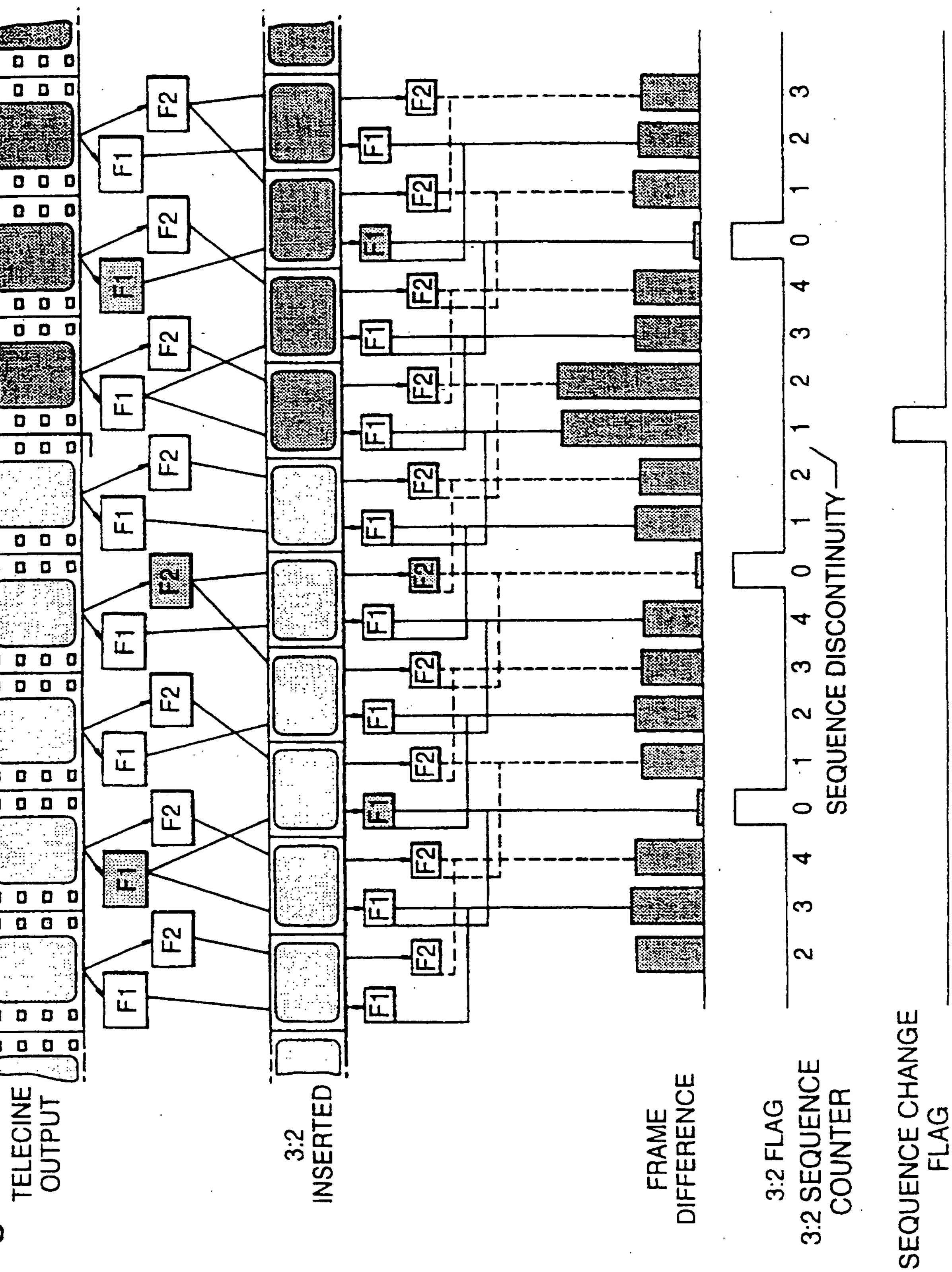

TELECINE OUTPUT
3:2 INSERTED
FRAME DIFFERENCE
3:2 FLAG
3:2 SEQUENCE COUNTER
F1
F2
2 3 4 0 1 2 3 4 0 1 2 3 4 0 1 2 3 4 0

Fig.5.

Fig.6.


TELECINE OUTPUT
3:2 INSERTED
FRAME DIFFERENCE
3:2 FLAG
3:2 SEQUENCE COUNTER
SEQUENCE CHANGE FLAG
EDIT POINT
SEQUENCE DISCONTINUITY
F1
F2
2 3 4 0 1 2 3 4 0 1 2 1 2 3 4 0 1 2 3

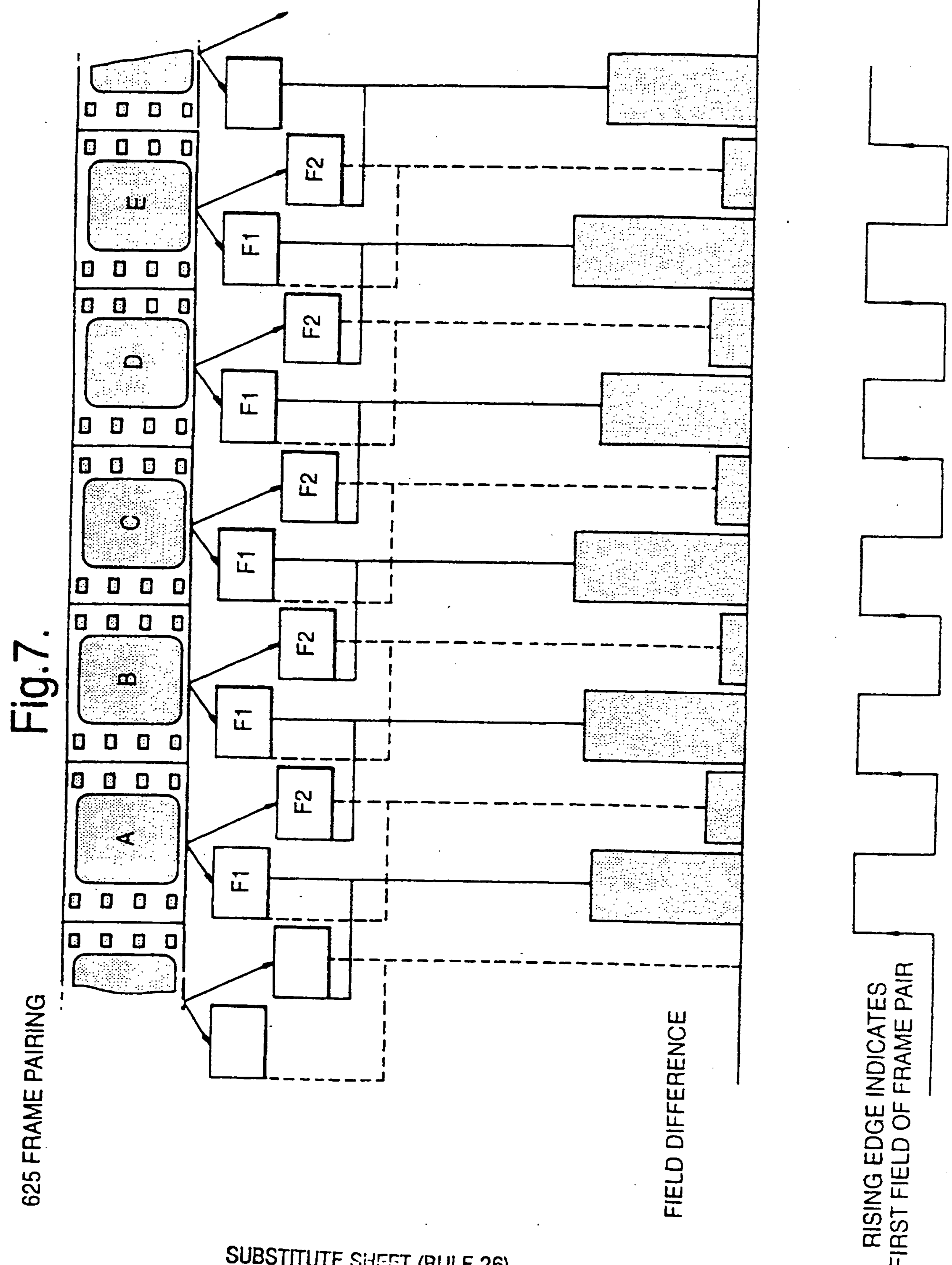
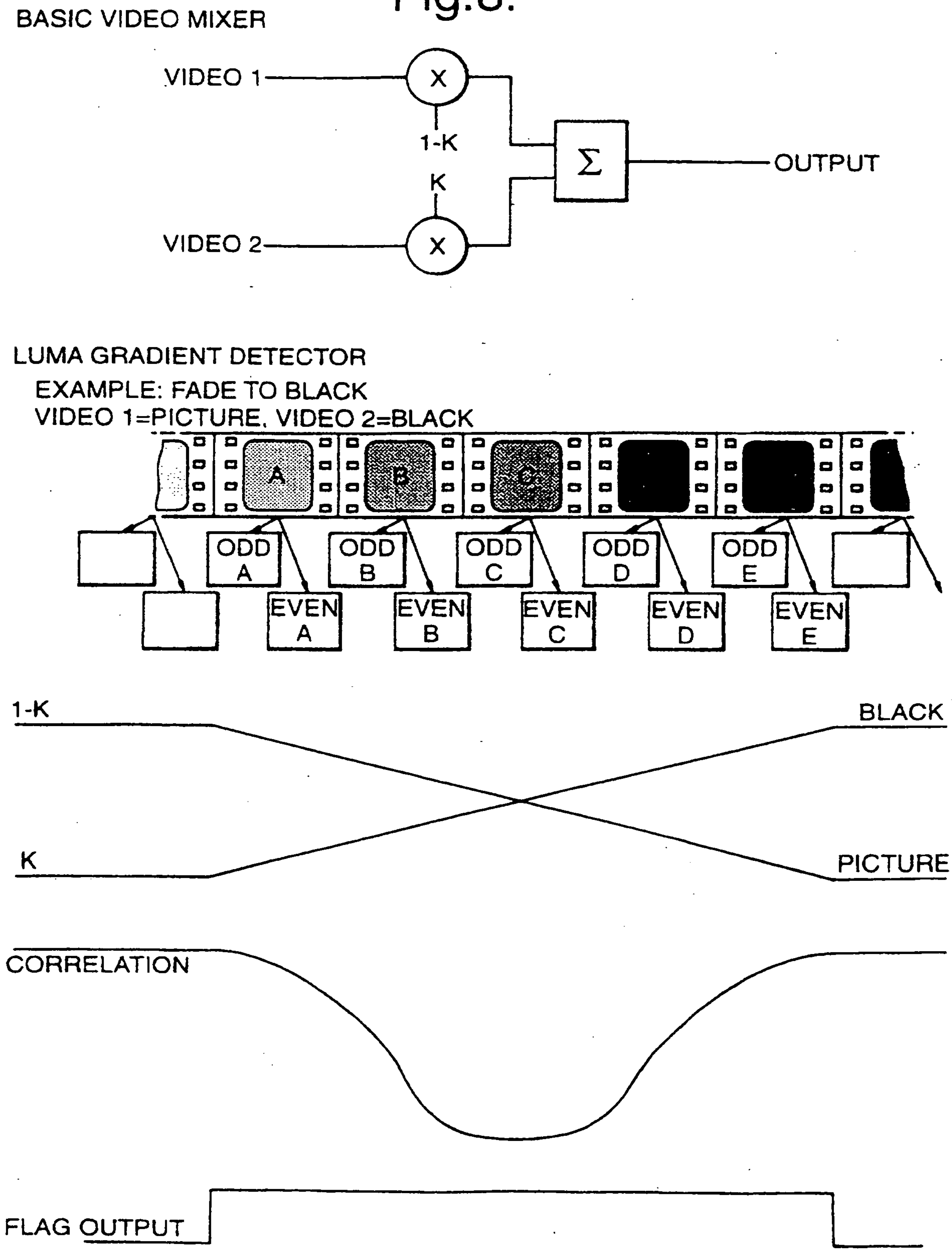

625 FRAME PAIRING
A
B
C
D
E
F1
F2
FIELD DIFFERENCE
RISING EDGE INDICATES FIRST FIELD OF FRAME PAIR

Fig.7.

Fig.8.


BASIC VIDEO MIXER
VIDEO 1
X
1-K
K
Σ
OUTPUT
VIDEO 2
X
LUMA GRADIENT DETECTOR
EXAMPLE: FADE TO BLACK
VIDEO 1=PICTURE, VIDEO 2=BLACK
A
B
C
ODD A
ODD B
ODD C
ODD D
ODD E
EVEN A
EVEN B
EVEN C
EVEN D
EVEN E
1-K
BLACK
K
PICTURE
CORRELATION
FLAG OUTPUT